

Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/105057/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Vaughan, Ian ORCID: <https://orcid.org/0000-0002-7263-3822>, Gotelli, Nicholas, Memmott, Jane, Pearson, Caitlin, Woodward, Guy and Symondson, William ORCID: <https://orcid.org/0000-0002-3343-4679> 2018. econullnetr: an R package using null models to analyse the structure of 1 ecological networks and identify resource selection. *Methods in Ecology and Evolution* 9 (3) , pp. 728-733. 10.1111/2041-210X.12907 file

Publishers page: <http://dx.doi.org/10.1111/2041-210X.12907>
<<http://dx.doi.org/10.1111/2041-210X.12907>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies.

See

<http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



`econullnetr`: an R package using null models to analyse the structure of
ecological networks and identify resource selection

Ian P. Vaughan^{1*}, Nicholas J. Gotelli², Jane Memmott³, Caitlin E. Pearson¹, Guy Woodward⁴,
William O.C. Symondson¹.

¹Cardiff School of Biosciences, Cardiff University, Cardiff, CF10 3AX, UK

²Department of Biology, University of Vermont, Burlington, VT 05405, USA

³Life Sciences Building, University of Bristol, Bristol, BS8 1TQ, UK

⁴Department of Life Sciences, Imperial College London, Silwood Park Campus, Buckhurst Road, Ascot,
Berkshire SL5 7PY, UK.

*Corresponding author: Email vaughanip@cardiff.ac.uk

Running title: `econullnetr`: null models for ecological networks

Word count: 3490 (incl. references, tables and figure caption); 2242 (main text)

Summary

1. Network analysis is increasingly widespread in ecology, with frequent questions asking which nodes (typically species) interact with one another and how strong are the interactions. Null models are a way of addressing these questions, helping to distinguish patterns driven by neutral mechanisms or sampling effects (e.g. relative abundance of different taxa, sampling completeness) from deterministic biological mechanisms (e.g. resource selection and avoidance), but few ‘off the shelf’ tools are available.
2. We present `econullnetr`, an R package combining null modelling and plotting functions for networks, with data-export tools to facilitate its use alongside existing network analysis packages. It models resource choices made by individual consumers, enabling it to capture individual-level heterogeneity and generalising to a wider range of data types and scenarios than models applied directly to network matrices. The outputs can be analysed from the level of individual links to whole networks.
3. We describe the main functions and provide two short examples, along with the results of a benchmarking exercise to provide guidance about the statistical power and error rates. Our hope is that `econullnetr` provides a basis for more widespread use of null modelling to assist ecological network interpretation.

Key words: food webs; plant-pollinator networks; prey choice; resource selection

Introduction

Network analysis is being used with increasing frequency in ecology, primarily for the analysis of food webs, and mutualistic and social networks. A common aim is to understand patterns of interactions – among nodes usually defined as species – and their strength – often approximated by interaction frequency – as this allows deeper insight into network assembly, functioning and response to disturbance (Allesina *et al.*, 2008). Related research questions ask whether particular species are specialists or generalists, and which taxa they interact with most strongly (e.g. Vázquez & Aizen, 2003).

In addressing these research themes with empirical data, null models can highlight network structures that cannot be accounted for by neutral mechanisms or sampling artefacts. A simple null hypothesis is that interaction frequencies between consumers and resources are a consequence of the relative abundances of the potential resources i.e. more common resources are utilised more heavily than rarer resources because they are encountered more often (Agusti *et al.*, 2003). This neutral mechanism has been applied in a series of studies, affording valuable insights into network structure and the behaviour of consumers (e.g. Agusti *et al.*, 2003; Vázquez & Aizen, 2003; Ibanez 2012). One of the main advantages is that, by holding characteristics of the observed data constant (e.g. sample size), null models can reduce the influence of sampling effects on network interpretation (Blüthgen *et al.*, 2008).

Here we present an R package, *econullnetr*, comprising functions for null modelling and interpretation of a wide range of networks, including bipartite and more complex networks. The model was developed to identify prey choices by predators (Agusti *et al.* 2003; King *et al.*, 2010; Davey *et al.*, 2013), but generalises to any network with one or more consumer

species. By operating at the level of individual consumers, rather than data already summarised at the node level in network matrices (e.g. Dormann *et al.*, 2008), it can account for additional sources of heterogeneity in the data (e.g. intra-specific heterogeneity) or the need to combine spatially and/or temporally distinct samples into an overall network. Furthermore, this individual-level approach generalises to complex interaction data, where: i) individual consumers may interact with multiple resources, ii) data could range from a list of species an individual interacted with, to actual quantities (e.g. number of prey eaten by a predator), and iii) “forbidden links” may need to be specified. Rather than fixing network properties (e.g. overall network size) *a priori*, network structure emerges from individual behaviour, which can lead to more robust models (Grimm & Berger, 2016).

We describe the null model and report on a benchmarking exercise to provide guidance about its power and error rate (Gotelli & Ulrich, 2012). We then describe the functions in `econullnetr` and provide two short examples. For simplicity, and in line with long-established conventions in ecological network analysis, we refer to the nodes as “species”, each of which comprises one or more individuals: in reality, nodes vary in their taxonomic resolution or may represent different entities altogether (e.g. functional groups), and `econullnetr` can be applied in these cases too.

Overview of the null model

Resource selection is modelled for each individual consumer in turn (see below), after which interaction strengths are summarised for each pair of consumer and resource species to produce the interaction matrix, from which a range of statistics can be calculated. Iterations of the model build sampling distributions for the selected statistics, from which $1-\alpha\%$ confidence limits can be calculated using the $1-\alpha/2$ percentiles from the frequency

distribution (Manly 2006). If the value of the statistic from the observed network falls outside the confidence interval, it can be considered to be significantly greater or less than expected based on random resource use (i.e. resources used in proportion to their abundance). The mean value for the statistic across the iterations of the model provides an estimate of the expected value to which the observed value can be compared.

Interaction data may be nominal or quantitative at the level of individual consumers, prior to being aggregated to species level in the interaction matrix (Table 1). To model resource selection with nominal interaction data, the same number of resource species are selected as were originally recorded (i.e. an individual's degree is held constant), using sampling without replacement: the probabilities of different resources being selected are proportional to their relative abundance (Agusti *et al.*, 2003). Direct measures of resource availability (e.g. flower abundance in quadrats or transects) are required by `econullnetr` (cf. using marginal totals of a network matrix e.g. Dormann 2009). For count data, the same number of interactions are allocated (e.g. prey items eaten or flowers visited by an individual), allowing repeated use of the same resource species (i.e. sampling with replacement). When interactions are quantities such as biomass, the proportions are drawn from a Dirichlet distribution with shape parameters determined by the relative abundance of the different resources (Ainsworth *et al.*, 2010). For both types of quantitative data (Table 1), the degree can either be held constant within each individual or resources drawn from the full range of species.

An important stage in null model development is investigating Type I and II error rates, respectively the frequency of erroneously identifying patterns in random data and failing to identify features in the data, using synthetic data with known properties (Gotelli & Ulrich, 2012). We subjected `econullnetr` to >100 benchmark tests capturing a range of sample

sizes (10–100 individuals per consumer species), data types (nominal and quantitative) and strengths of resource selection by consumers, the latter ranging from no selection (‘perfect generalists’) to only interacting with one preferred resource (‘perfect specialists’). Performance was examined for individual links, whole species and the complete network. Full details of the testing process and the results are provided in Appendix S1. In summary, Type I error rates were consistently around 5%, as expected for 95% confidence limits, except for the combination of multinomial data with three out of the five network-level statistics (linkage density, weighted connectance and interaction evenness), for which Type I error >10% (Table S2). No problems were evident at link- or species-level (Table S2). The ability to detect preferences was strongly related to the strength of a consumer’s preferences and secondarily to sample size: very strong preferences (e.g. focusing on a single resource species) were detected in almost all situations, even at small sample sizes, whilst weak preferences were hard to detect, irrespective of sample size (Tables S3 & S4; Figs S2 & S3). The tests provide confidence that strong patterns of resource preference should be detected and results not compromised by frequent spurious test results.

The `econullnetr` package

The package comprises six functions and three data sets (Table 2). We briefly illustrate some of the features using the *Silene* (flower visitation) and Broadstone (food web) data sets: `econullnetr`’s help files and vignette provide full details and code to reproduce the examples.

The main function, `generate_null_net`, uses two data frames and an optional third: i) interaction data, where rows represent individual consumers, columns the resource species and the elements either the presence/absence of an interaction, the number of interactions or

the quantity; ii) the relative abundance of the resource species; and iii) optionally, a table of resource weights, one set for each consumer species, which are multiplied by the relative resource abundances to modulate the probabilities of different resources species being selected. This is primarily to specify ‘forbidden’ links for each consumer (weight = 0). Where data from different time points or locations are combined to produce an overall network, this can be specified in the call to `generate_null_net`, running the model separately in each sub-unit before combining them.

The Silene network

For bipartite networks, a wide range of statistics can be calculated for the observed and null networks using `bipartite_stats` (Table 2). *Silene* shows significantly lower linkage density, connectance and interaction evenness than expected (Table 3), consistent with preferences among the pollinator community. For any network, `test_interactions` generates a table comparing the observed and expected link strengths for all interactions in the network. Only 6% of links in *Silene* were inconsistent with the null model (Fig 1), including preferences for *Senecio* by two of the commonest pollinators (*Eristalis tenax* and *Rhagonycha fulva*), and fewer visits than expected to *Hypericum* by *Episyrphus balteatus* (Fig 1).

Whilst network plots are valuable for revealing where preferences occur, they are less effective at conveying the strength of preferences: instead, `plot_preferences` provides a detailed comparison of observed and predicted link strengths for individual consumer taxa (King *et al.*, 2010). One pollinator, the hoverfly *Sphaerophoria scripta*, showed a near four-fold preference for *Silene* at the expense of *Senecio*, which was only visited around 20% as often as predicted: visitation to the other flower species was consistent with the null model

(Fig 2). In a conservation context, the apparent specialisation of this hoverfly, coupled with its abundance, highlights its potential importance for managing *Silene* (Gibson *et al.* 2006).

The Broadstone Stream network

For more complex networks, a range of network packages, such as `igraph` (Csardi & Nepusz 2006) or `cheddar` (Hudson *et al.*, 2013), may be useful for calculating network statistics and visualising the data. The function `generate_edgelist` exports the observed and expected link strengths in a format that is readily imported into other packages. Here, we used the `cheddar` package in combination with `econullnetr`.

Eighteen percent of links in the Broadstone food web were inconsistent with the null model (Fig 3 a & b), with all but one of the predators displaying evidence of prey preferences (positive or negative; Fig 3a). Combining `cheddar`'s `NodeQuantitativeDescriptors` function with `generate_null_net`, revealed one predator (*Cordulegaster*) showed higher generality than expected, two (*Trissopelopia* and *Macropelopia*) demonstrated dietary specialisation, whilst the remainder were consistent with the null model (Fig 3b). The preferences fitted the well-known size structuring of the Broadstone food web (Woodward *et al.* 2005). *Cordulegaster*, the largest invertebrate predator, consumed small, abundant taxa less often than expected and larger, less abundant taxa more often than expected, leading to its more general diet than predicted by the model (Fig 3c). This reflects mouthparts that allow it to take larger prey and the relative inaccessibility of small prey in fine interstitial spaces. *Trissopelopia* (*Macropelopia* was similar, but not shown) was one of the smallest predators, and showed strong preference for *Heterotrissocladius*, a small chironomid midge that occupies similar interstitial microhabitats, with most other taxa consistent with the null model or eaten less often than

expected (e.g. *Zavreliomyia*), leading to the more specialised diet (Fig 3d). Large, mobile taxa (towards the bottom of the y-axis) were rarely consumed by this small and relatively sedentary predator.

Concluding remarks

The `econullnetr` package provides simple tools for revealing structures within networks. It ignores dynamics, only considering networks at a point in time, and uses simple proxies of interaction strength (e.g. interaction frequency) that may limit insight into network behaviour (Berlow *et al.*, 2004). Nevertheless, `econullnetr` covers the types of data most frequently collected, and should aid basic network interpretation. The model cannot explain the mechanisms underpinning ‘selection’, which may be an active choice (e.g. based on prey size or flower morphology) or a passive process (e.g. spatio-temporal separation), but provides a way to highlight interactions for further investigation. It has good Type I error properties under a wide range of conditions and, with sufficient consumers sampled ($>>10$ individuals), good power to recover resource preferences.

As networks become large and contain many links, multiple significance testing is likely to become important. Currently, `econullnetr` does not attempt to control for this, but issues a reminder about Type I error when `test_interactions` is run. Based on our benchmark testing, Type I error rates were low and could be reduced by selecting a more stringent testing criterion (e.g. $\alpha=0.01$). For larger networks, some form of false discovery rate procedure is likely to be valuable for controlling Type I errors (e.g. the local false discovery rate, Gotelli & Ulrich 2010), and we hope to add this in a future version.

Acknowledgements

We would like to thank David Gilljam for formatting the Broadstone Stream data, Rose Moorhouse-Gann for helping to test the package, and Igansi Bartomeus and an anonymous reviewer for helpful comments on an earlier version of the manuscript. Development of `econullnetr` was supported by a Cardiff University Research Leave Fellowship awarded to IPV.

Accessibility

`econullnetr` is freely available under the open source MIT licence. A stable version is available from the CRAN archive (<https://CRAN.R-project.org/package=econullnetr>), whilst the development version is hosted on GitHub (<https://github.com/ivaughan/econullnetr>), where feature requests and bug reports can be posted.

Author contributions

IPV designed the null model, wrote the R package and the first draft of the paper. NJG helped to design the benchmarking tests and advised on aspects of the null model. JM, CEP and GW contributed data sets and ideas on model applications. WOCS provided applications for the initial development of the model and helped to design the plotting functions. All authors were involved in preparing the final manuscript.

References

Agustí, N., Shayler, S.P., Harwood, J.D., Vaughan, I.P., Sunderland, K.D. & Symondson, W.O.C. (2003) Collembola as alternative prey sustaining spiders in arable ecosystems: prey detection within predators using molecular markers. *Molecular Ecology*, **12**, 3467–3475.

239 Ainsworth, C.H., Kaplan, I.C., Levin, P.S. Mangel, M. (2010) A statistical approach for
 240 estimating fish diet compositions from multiple data sources: Gulf of California case study.
 241 *Ecological Applications*, **20**, 2188–2202.

242

243 Allesina, S., Alonso, D. & Pascual, M. (2008) A general model for food web structure.
 244 *Science*, **320**, 658–661.

245

246 Berlow, E.L., Neutel, A.-M., Cohen, J.E., De Ruiter, P.C., Ebenman, B., Emmerson, M., Fox,
 247 J.W., Jansen, V.A.A., Jones, J.I., Kokkoris, G.D., Logofet, D.O., McKane, A.J., Montoya,
 248 J.M. & Petchey, O. (2004) Interaction strengths in food webs: issues and opportunities.
 249 *Journal of Animal Ecology*, **73**, 585–598

250

251 Blüthgen, N., Fründ, J., Vázquez, D.P. & Menzel, F. (2008) What do interaction network
 252 metrics tell us about specialization and biological traits? *Ecology*, **89**, 3387–3399.

253

254 Csardi, G. & Nepusz, T. (2006) The igraph software package for complex network research,
 255 *InterJournal, Complex Systems*, **1695**.

256

257 Davey, J.S., Vaughan, I.P., King, R.A., Bell, J.R., Bohan, D.A., Bruford, M.W., Holland,
 258 J.M. & Symondson, W.O.C. (2013) Intraguild predation in winter wheat: prey choice by a
 259 common epigeal carabid consuming spiders. *Journal of Applied Ecology*, **50**, 271–279.

260

261 Dormann, C.F., Gruber B. & Fründ, J. (2008). Introducing the bipartite package: analysing
 262 ecological networks. *R news*, **8**, 8–11.

263

Dormann, C.F., Fründ, J., Blüthgen, N. & Gruber, B. (2009). Indices, graphs and null models: analyzing bipartite ecological networks. *Open Ecology Journal* **2**, 7–24.

Gibson, R.H., Nelson, I.L., Hopkins, G.W., Hamlett, B.J. & Memmott, J. (2006) Pollinator webs, plant communities and the conservation of rare plants: arable weeds as a case study. *Journal of Applied Ecology*, **43**, 246–257

Gotelli, N.J. & Ulrich, W. (2010) The empirical Bayes approach as a tool to identify non-random species associations. *Oecologia*, **162**, 463–477.

Gotelli, N.J. & Ulrich, W. (2012) Statistical challenges in null model analysis. *Oikos*, **121**, 171–180.

Grimm, V. & Berger, U. (2016) Next-generation ecological modelling. *Ecological Modelling*, **326**, 1–3.

Hudson, L.N., Emerson, R., Jenkins, G.B., Layer, K., Ledger, M.E., Pichler, D.E., Thompson, M.S.A., O’Gorman, E.J., Woodward, G & Reuman, D.C. (2013) Cheddar: analysis and visualisation of ecological communities in R. *Methods in Ecology and Evolution*, **4**, 99–104.

Ibanez, S. (2012) Optimizing size thresholds in a plant–pollinator interaction web: towards a mechanistic understanding of ecological networks. *Oecologia*, **170**, 233–242.

288 King, R.A, Vaughan, I.P., Bell, J.R., Bohan, D.A, & Symondson, W.O.C. (2010) Prey choice
289 by carabid beetles feeding on an earthworm community analysed using species- and lineage-
290 specific PCR primers. *Molecular Ecology*, **19**, 1721–1732.

291

292 Manly, B.F.J. (2006) *Randomization, Bootstrap and Monte Carlo Methods in Biology* (3rd
293 edn). Chapman & Hall, Boca Raton.

294

295 Vázquez, D.P. & Aizen, M.A. (2003) Null model analyses of specialization in plant-
296 pollinator interactions. *Ecology*, **84**, 2493–2501.

297

298 Woodward, G., Speirs, D.C. & Hildrew, A.G. (2005) Quantification and resolution of a
299 complex, size-structured food web. *Advances in Ecological Research*, **36**, 84–135.

Table 1. The four types of interaction data handled by `econullnetr`.

Data recorded for individual consumers	Examples
<i>Nominal</i>	
One resource species recorded per consumer	<ul style="list-style-type: none">• Ants carrying individual seeds• Pollinators recorded on the flower species where they were observed
Varying numbers of resource species per consumer	<ul style="list-style-type: none">• List of flower species visited by a pollinator during a fixed observation period• Molecular gut contents analysis identifying prey species, but without quantifying the amount of prey tissue
<i>Quantitative</i>	
Counts of interactions with each resource species by individual consumers	<ul style="list-style-type: none">• Number of visits to each flower species by a pollinator during a fixed observation period• Visual gut contents analysis, counting the number of individuals of each prey species based on identification of hard parts
Biomass, or equivalent quantitative measurement	<ul style="list-style-type: none">• Proportion of gut contents comprised by different food sources• Number of pollen grains of different flower species on a pollinator's body

Table 2. Overview of `econullnetr`'s functions and data sets

Name	Description
<i>Functions</i>	
<code>generate_null_net</code>	Specify and run the null model.
<code>test_interactions</code>	Compare observed interaction strengths to those generated by the null model
<code>plot_preferences</code>	Plot observed and modelled interaction strengths for individual consumer species
<code>bipartite_stats</code>	Compare network metrics between the observed and null bipartite networks. A wrapper for the <code>bipartite</code> package's <code>networklevel</code> , <code>grouplevel</code> and <code>specieslevel</code> functions (Dormann <i>et al.</i> , 2008, 2009).
<code>plot_bipartite</code>	A wrapper for <code>bipartite</code> 's <code>plotweb</code> function (Dormann <i>et al.</i> , 2008), colour coding interactions according to their consistency with the null model
<code>generate_edgelist</code>	Export null model results in a format compatible with other network analysis packages
<i>Data sets</i>	
Broadstone	Part of the highly-resolved Broadstone Stream food web (Woodward <i>et al.</i> , 2005).
Silene	Flower visitation network, notable for the presence of small-flowered catchfly <i>Silene gallica</i> , a rare arable weed in the UK (Gibson <i>et al.</i> 2006)
WelshStreams	Part of a macroinvertebrate food web from upland streams in south Wales, UK, focusing on two abundant predators

Table 3. Network-level statistics for *Silene*, comparing observed values to the 95% confidence limits from the null model and including the standardised effect size (SES).

Metric	Observed	Null	Lower.CL	Upper.CL	Test	SES
linkage density	5.096	6.687	6.166	7.230	Lower	-5.830
Connectance (weighted)	0.142	0.186	0.171	0.201	Lower	-5.740
Nestedness (weighted)	0.518	0.569	0.369	0.744	ns	-0.535
Interaction evenness	0.849	0.897	0.880	0.913	Lower	-5.763

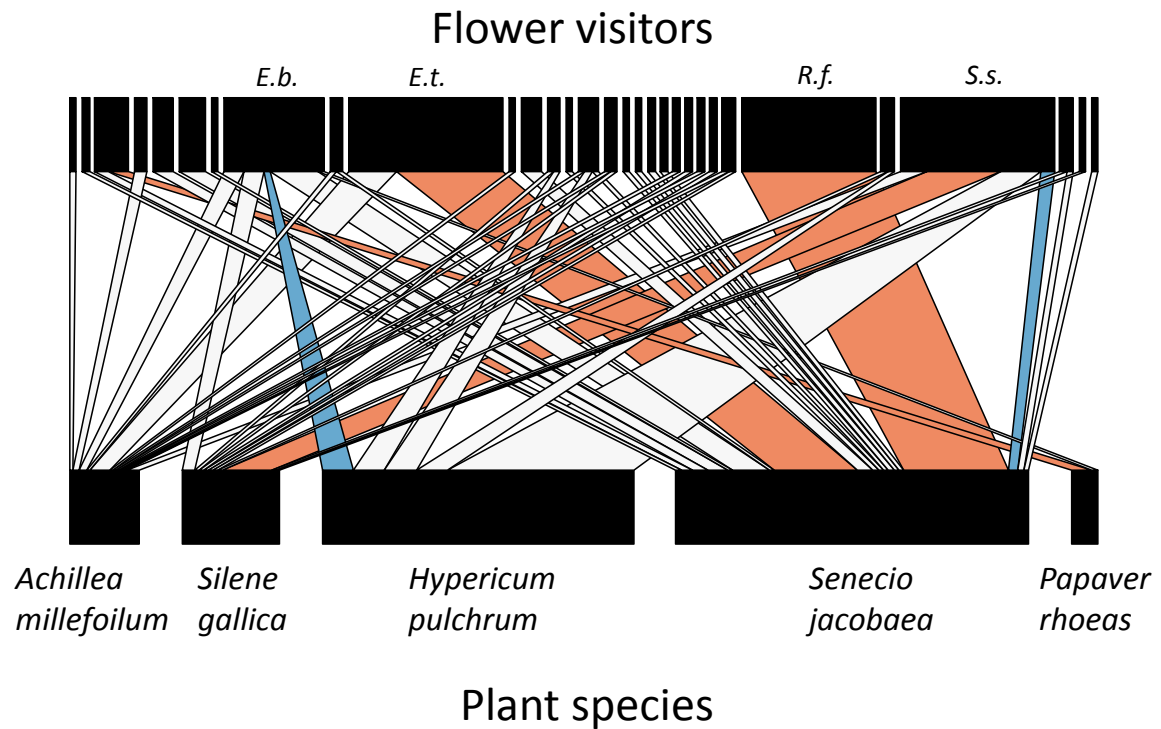


Figure 1. Flower visitation network for *Silene* using `plot_bipartite`: for simplicity, individual pollinator species are not labelled. Link widths represent the observed frequency of interactions, with red links stronger than expected under the null model and blue links weaker than expected. Bar widths at the two levels indicate the relative abundances of different plants and flower visitors. Four pollinator species mentioned in the test are highlighted: *E.b.* = *Episyrphus balteatus*, *E.t.* = *Eristalis tenax*, *R.f.* = *Rhagonycha fulva* and *S.s.* = *Sphaerophoria scripta*.

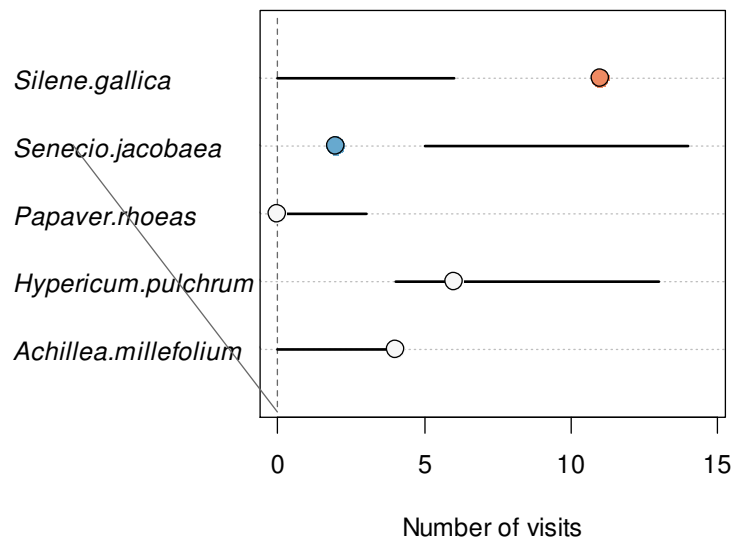


Figure 2. Preference plot for the hoverfly *Sphaerophoria scripta* in the *Silene* network, comparing the observed interaction frequencies (dots) to the 95% confidence intervals from the null model (bars). The red dot denotes an interaction that was stronger than expected under the null model, the blue dot weaker than expected, and white dots consistent with the null model.

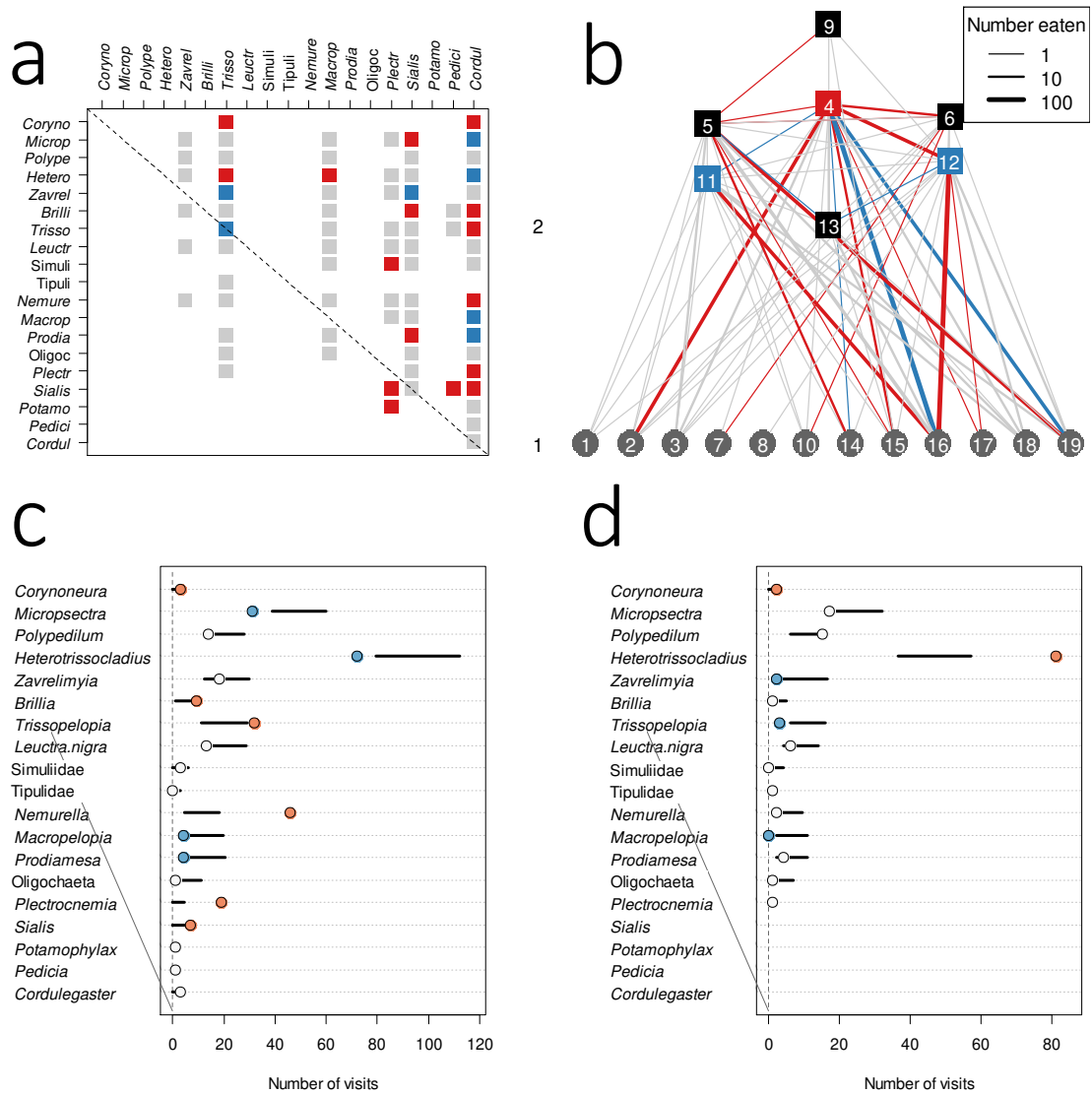


Figure 3. Null modelling results for Broadstone: (a) the predation matrix, with taxa ranked in increasing size order (left to right); grey symbols indicate links consistent with the null model, blue = significantly weaker, red = significantly stronger and white = interaction not observed; (b) alternative plot of the web, with the mean trophic level of each species on the y-axis; links colour coded as for (a), with thickness indicative of predation frequency; predators in black = consistent with the null model, red = more general diet than expected, blue = more specialised; (c) preferences of *Cordulegaster* (predator 4 in (b)) and (d) *Trissopelopia* (predator 12 in (b)); interpretation of (c) and (d) as for Figure 2.

300 **Supporting information**

301

302 **Appendix S1.** Full methods and results from benchmark testing of the null model.

303

304